知識データベースを応用したWeb取説検索システム

Web-Based Retrieval System for Product Operation Manuals Based on Knowledge Database Architecture

鎌田 喬浩, 梶 雅代, 堀内 直明
Takahiro Kamada, Masayo Kaji, Naoaki Horiuchi

要　旨　民生機器の多機能化に伴い，ユーザが手にする取扱説明書の情報量は増加し続けている。その結果，ユーザが所望の操作方法を探す際に，取扱説明書を容易に活用することが次第に難しくなってきている。そこで我々は，知識データベースを用いて取扱説明書の活用を支援する，Web上で閲覧可能な検索システム(Web取説検索システム)を開発した。本システムの知識データベースとは，取扱説明書の階層構造と特定の用語に対する多様なゆらぎ表現をデータベース化したものである。本システムを利用すれば，ユーザは思いつくキーワードを入力するだけで期待する取扱説明書の内容を素早く得ることができる。

Summary　Consumer electronic products have equipped much more functionalities than ever before, so that the volume of their operation manuals has kept on increasing, and that leads to a serious problem that it is becoming harder for users to utilize the manuals for acquiring information that they need. Therefore, we developed a Web-based retrieval system for product operation manuals taking advantage of knowledge database architecture. The knowledge database comprises a hierarchy structure of content description in the manuals and a variety of terms that had a similar meaning to some specific words included in the contents. We expected this system would allow users to get desirable information from the manuals more effectively, only by indicating keywords they could come up with easily.

キーワード ： 知識データベース，取扱説明書，情報検索，Web取説

1. はじめに

民生機器の操作方法とその説明は，一般に紙冊子の取扱説明書として提供されている。紙冊子の取扱説明書では，ユーザが操作方法の記述を容易に見つけられるように，以前から目次や索引といった検索性を高める手法が活用されてきた。ところが，近年のように民生機器の多機能化にともなって取扱説明書の情報量が増加すると，それに併せて目次や索引の記述も増加するため，これまでのような検索性の向上は期待できない。したがって，大量な情報を含む取扱説明書に対して，ユーザが所望する情報を容易に検索可能にする新たな手法が必要となっている。

一方，情報処理技術の分野では，大量の情報の中から所望する情報を容易に検索するための研究が進められている。特に，情報検索の支援技術として，知識データベースの研究が盛んである。その一例として，カーナビゲーションシステムにおける地点検索への応用が挙げられる[1]。このカーナビゲーションシステムにはキーワードから目的地を類推する辞書が用意されており，目的地の名称を正確に知らないユーザでも目的地設定が容易にできるシステムとなっている。

このように知識データベースを情報検索に応用すれば，キーワードからユーザの所望する情報を類推し，ユーザの要求を考慮した検索支援を行える。

本稿は，Web上で閲覧や検索が可能な取扱説明書の情報検索システムに対し，知識データベースを応用

した事例について解説する。先ず，従来の検索手法を取扱説明書に適用した場合の課題を明らかにし，次にその解決手段である知識データベースの導入について述べる。その後，上記知識データベースの具体的な構築方法とそれを利用した検索アルゴリズムについて解説を行い，最後に今回開発したWeb取説検索システムの概要を紹介する。

2. 取扱説明書の検索における課題

ドキュメントの集合に対してキーワード検索を行うシステムは，以前から存在している。その多くは，ユーザが入力したキーワードと一致する文字列をドキュメント内から検索し，検索結果として同文字列の記述箇所を明示するものである。

従来のシステムを取扱説明書に適用した場合には，以下の２つの課題が挙げられる。

●課題１：文書の階層構造が反映されない

民生機器の多くは，基本的な機能だけはなく，ユーザが必要に応じて利用できる多くの関連する機能を有している。そして，取扱説明書の記述も通常の文章とは異なり，機能ごとに説明文が分類され，階層的に記述されている。そのため，検索結果としては機能ごとの説明文が目的に則して並列に，あるいは階層的に提供されることが，ユーザが機器を理解し使いこなす上で望ましい。しかし，従来のシステムでは取扱説明書のすべての記述を連続した文字列として扱うため，取扱説明書の持つ特有の構造や，機能ごとの区切りを検索結果に反映することができない。その結果，ユーザは検索結果と目的とする操作との一致関係を把握することができず，機器操作についての適切な情報を得ることができない。

●課題２：専門知識を必要とする

民生機器用に作成された取扱説明書には，メーカー特有の機能名称や専門用語が多く含まれている。

そのため，ユーザは所望の情報を得るために，取扱説明書で用いられている特有の用語を推測し，あるいは一般的な用語でも取扱説明書で記述された正しいつづりを知った上で検索キーワードを決定し，システムに入力する必要がある。しかし，機器の知識が乏しいユーザにとっては，取扱説明書の記述を正確に推測することはきわめて難しいため，適切なキーワードを選べないと，検索洩れや誤った検索結果を誘発する恐れがある。

3. 知識データベースの導入

我々は上記の課題を解決するために，以下に解説する「構造化した取扱説明書」「キーワードの重要度」「ゆらぎ辞書」を知識データベースとして構築した。

●構造化した取扱説明書データベース

検索対象として，取扱説明書の記述内容を個別の機能や説明項目ごとに分類し，それらの階層構造に沿って構造化したデータを用いる。その結果，項目単位の検索が可能となり，検索結果に該当する項目を，ユーザの目的に則して並列，あるいは階層的に列挙することができる。

●キーワードの重要度データベース

さらに，構造化した取扱説明書の内容を言語的に解析し，各項目の特徴をそこに含まれるキーワードの重要度として数値化する。これによって，意味内容に基づいた検索結果の順位付けやソート，さらにはユーザが所望する情報に該当する項目を優先的に表示することが可能になると考えられる。

●ユーザ入力支援のためのゆらぎ辞書

ユーザが入力したキーワードを取扱説明書に記述されている適切な文字列に変換し，所望の検索結果を出力させる。具体的には，ユーザが入力するキーワードと，取扱説明書の記述との対応関係を対応させたゆらぎ辞書を利用し，入力されたキーワードが取扱説明書に存在しない場合に限って変換処理を行う。これによって，ユーザが取扱説明書に未記載のキーワードを入力した場合であっても当初の目的に近い内容の検索が可能となる。つまり，ゆらぎ辞書を用意することによって，キーワード入力に必要とされるユーザの知識量を軽減する効果があると考えられる。

各知識データベースの具体的な構築方法については，次の4章にて詳しく解説する。

4. 知識データベースの構築方法

4.1 取扱説明書の構造化

本システムでは，紙冊子として配布するために作成された民生機器の取扱説明書 (注1) のデータを用いた。データの一部を**図1**に示す。

図1に示すとおり，取扱説明書のデータは文章と図表から構成されており，それぞれの情報には，体系化された装飾やレイアウトがなされていることが分かる。これは編者が，記述する情報に対して，その意味内容をもとに分類，整理したことを表している。そこで，本システムでは，そのような表記法則を情報の意

味体系として見なし，取扱説明書の構造化を進めた。

今回用いた取扱説明書では，設定方法，操作方法を解説する際に，先ず課題を掲げその解決方法を述べる形式をとっている。本稿では，一つの課題とその解決方法について記述された情報の固まりを，“トピック”と呼ぶものとする。

もし課題の概念が大きく，複数の課題に分割可能な場合は，階層構造を用いることで，トピック間の意味体系が表現されている。

例えば図１では，最上部に「番組の詳細を表示する」という概念の大きい課題が示されている。この課題は，「チャンネル情報を表示する」「番組情報を表示する」の，２つの具体的な課題に分割される。また，「チャンネル情報を表示する」課題に付随する課題として，「チャンネル情報の表示サイズを切り換えるには」が記されている。

このように，図１の記述内容を要素化し，論理関係を反映して構造化すると，**図２**に示す各トピックの階層構造で記述することができる。図２では，四角形が各トピックを表しており，最上部に課題，その下にトピック内の情報の要素を表している。一方，四角形を結ぶ線分は，トピック間の概念的な並列関係あるいは階層構造を表している。

以上の方法を用いて，取扱説明書に記述されてい
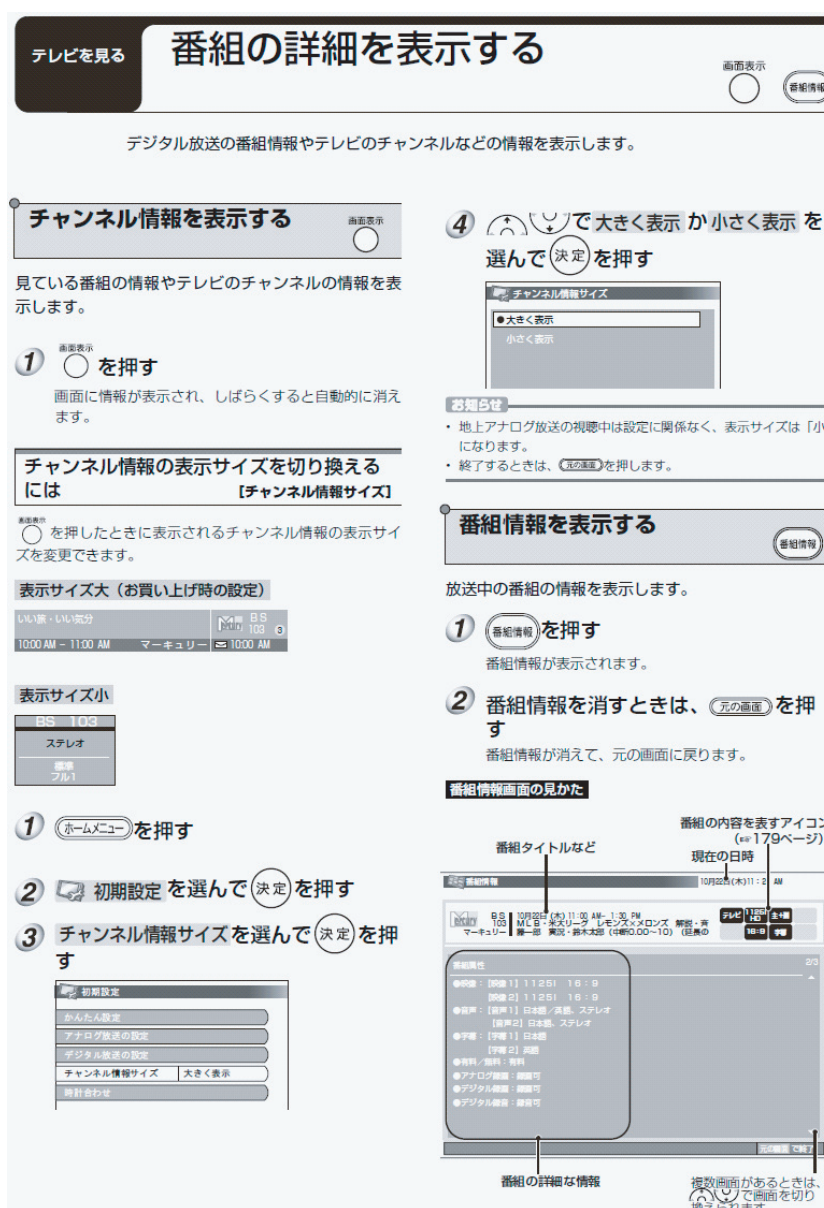
図１　取扱説明書の記述例

Fig 1.　A sample format of the operation manual description

番組の詳細を表示する
要旨

チャンネルの情報を表示する
要旨
手順 1

チャンネルの情報の表示サイズを切り換える
要旨
キャプション
図
キャプション
図
手順 1
手順 2
手順 3
手順 4
お知らせ

番組の情報を表示する
要旨
手順 1
手順 2
キャプション
図
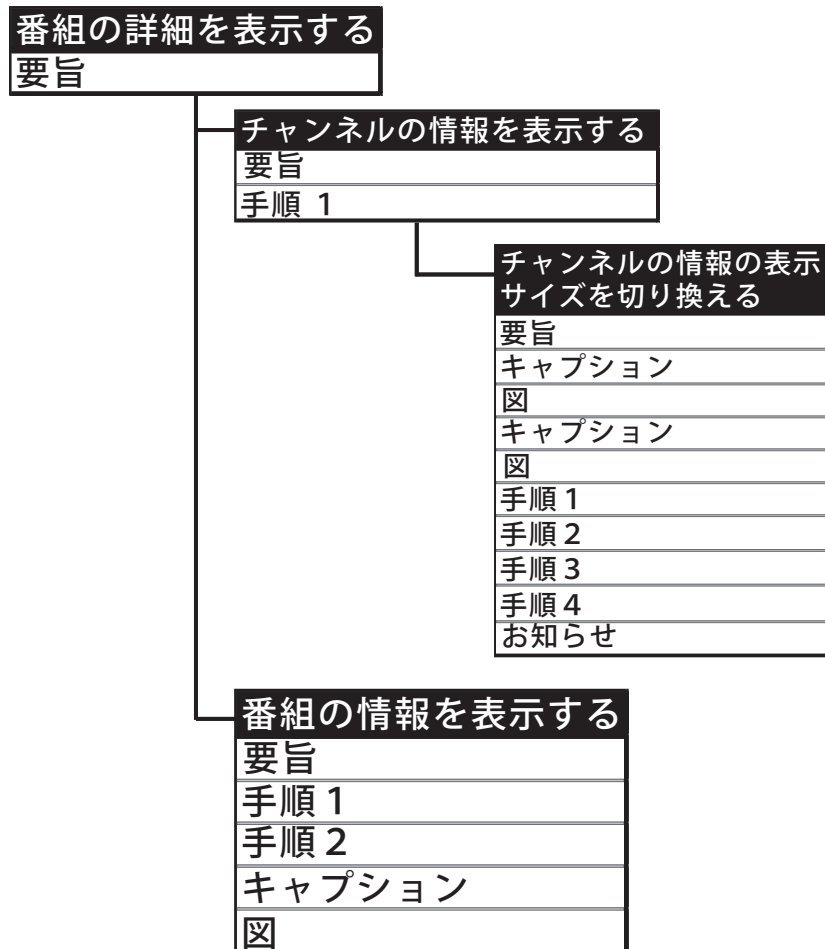
図 2　構造化した取扱説明書

Fig. 2　Hierarchy structure of topics

る全ての情報の構造化を行い，関連するすべてのトピック構造を知識データベースとして統合した。

4.2　重要度データベースの構築

キーワード検索の結果，複数のトピックが該当した場合は，ユーザが所望する情報を優先的に表示できることが望ましい。これを実現するには，ユーザの所望する情報が該当したトピックにどの程度含まれているかを推定し，その結果に従って該当項目を並び替える手法が考えられる。

そこで本システムでは，ユーザの所望する情報とトピックの内容との一致度合いを推定する指標として，“トピックにおけるキーワードの重要度”という概念を導入した。なぜなら，キーワードはユーザの要求を代表的に表現したものであるため，ある特定のトピックでそれらのキーワードが重要な位置付けを占める程，同トピックが所望する情報に近いと推定できるためである。

上記重要度の算出には TF-IDF 法を利用した。TF-IDF は式 1 で求められる数値であり，文章中のある単語の出現頻度 (Term Frequency：tf) と，検索対象とした全ての文章数 (N) における，ある単語が出現する文章数の割合の逆数 (Inverse Document Frequency：idf) の積より得られる。

$$TF-IDF = tf \times \log_{10}\left(\frac{N}{df}\right) \quad (式 1)$$

もし，あるキーワードが特定のトピック A での出現頻度が高く，かつ他のトピックでの出現頻度が低ければ，TF-IDF 値が高くなり，トピック A における同キーワードの重要度は高いと判断される。一方，同キーワードが他の多くのトピックに広く出現していた場合には「ありふれたキーワード」として TF-IDF 値が低くなり，トピック A における同キーワードの重要度は低いと判断される。

例えば，ユーザが指定したキーワードから複数の検索結果が得られた場合には，少なくとも一つのキーワードが複数のトピック中に出現していることを示している。この場合には，TF-IDF 値という重要度の概念から，同キーワードの重要度が高いトピックを優先的に表示させることができる。

本システムでは，取扱説明書の内容が動的には変化しない仮定のもと，あらかじめ全トピックに出現する全ての単語の TF-IDF 値を算出し，重要度データベースを構築した。最大の理由は，演算処理の多い重要度の算出を，検索時に実行せずに済むため，検索処理時間を短縮できることである。

4.3　ゆらぎ辞書の構築

ユーザが検索キーワードとして入力した文字列が，取扱説明書に記述されている文字列と表記が一致しない場合，両者の意図するものは同じであっても，単純な文字列の一致検索では検出できない。

本稿では，意図する物事が同一であってもユーザによって異なる表記として現れる傾向，すなわち検索キーワードのさまざまな変化を「ゆらぎ」と定義する。検索ワードのゆらぎは，以下に示す「表記ゆらぎ」,「同義語」,「類義語」の３種に大別可能である。以下に実例を示す。

【表記ゆらぎ】

全角 - 半角：コンピューター　⇔ ｺﾝﾋﾟｭｰﾀｰ

大文字 - 小文字：COMPUTER　⇔ computer

送り仮名：取り扱い ⇔　取扱い

【同義語】

略語：パーソナルコンピュータ ⇔ パソコン

略語：パーソナルコンピュータ ⇔ PC

【類義語】

パーソナルコンピュータ ⇔ ノートパソコン

次に，検索キーワードのゆらぎを解消し，それを取扱説明書の記述に変換するために必要な知識データベースである「ゆらぎ辞書」の構築方法について述べる。ゆらぎ辞書を構築するには以下に示す３段階の手順を踏む。

手順 1：待ち受け語の選定

取扱説明書に出現する単語の中から，ゆらぎが生じると考えられる「待ち受け語」を選定する。

手順 2：ゆらぎ語の作成

待ち受け語に対して，ゆらぎを生じさせた「ゆらぎ語」を作成する。

手順 3：データベースの構築

待ち受け語とゆらぎ語の対応関係を反映した形式で，データベースを構築する。

なお，ゆらぎ語の作成において,「全角 - 半角ゆらぎ」,「大文字 - 小文字ゆらぎ」については，そのゆらぎ方に法則性が見られるため，ゆらぎ語を自動的に作成した。

5.　知識データベースを用いた検索アルゴリズム

5.1　簡易自然文処理による検索キーワードの生成

情報検索システムとしてより多くのユーザにとって利用し易い形態の一つが，自然文入力である。つまり，スペースで区切られたキーワードの入力をユーザに強制するのではなく，自然な文章で自由に入力させるのである。しかし，一般に任意の自然文から検索キーワードを抽出する処理 (形態素解析処理など) は演算処理が重く，また自然文というユーザの自由度が高い内容から抽出したキーワードの場合，その信頼性の判断が難しいという課題がある。そこで本システムでは，取扱説明書に特化した簡易的な自然文処理を施し，キーワード入力を前提とした検索を可能とした。取扱説明書において，キーワードに成り得る重要な語句は，漢字，カタカナ，英数字によって表記されるものが大半を占めている。つまり，自然文で入力されても，平仮名や句読点は，有効なキーワードに成り得ない。そこで，平仮名と句読点の登場箇所を，キーワードを分割する位置として認識することとした。

以下に自然文からキーワードを抽出する例を示す。

【自然文入力】

デジタルカメラの画像を表示したい。

【分割位置の認識】

デジタルカメラ (の) 画像 (を) 表示 (したい。)

【抽出結果】

デジタルカメラ，　画像，　表示

例では，自然文をキーワードに分割する目印として認識した，平仮名と句読点の出現位置を括弧で示している。その出現位置に従って自然文を分割すると，デジタルカメラ , 画像 , 表示の３個のキーワードが抽出される。

勿論，平仮名のキーワードも取扱説明書に存在しないわけではない。そのため，簡易自然文処理は，通常のキーワード検索を実行した結果，ヒット項目が存在しない場合に限定して作動することとした。

5.2　検索アルゴリズム

本システムの検索アルゴリズムのフローチャート

を図3に示し，記述に沿って解説する。

(検索手順)

手順1．ゆらぎ辞書を参照し，入力単語群にゆらぎ語として，該当するものが無いか検索する。

手順2．手順1の結果，ゆらぎ語が存在する場合，待ち受け語に変換する。

手順3．取扱説明書データベースを参照し，キーワードが含まれるトピックを検索する。

手順4．手順3の結果，該当するトピックが存在する場合,手順7に移行する。存在しない場合は,手順5に移行する。

手順5．一度簡易自然文処理を行った場合は，検索アルゴリズムを終了する。実施していない場合は手順6に移行する。

手順6．簡易自然文処理を実行し，検索ワードからキーワードを抽出する。手順1に移行する。

手順7．取扱説明書の情報を抽出し，処理を終了する。

次に本システムを用いた，検索の具体例を示す。例では，ユーザが検索ワードとして「地デジのEPGを表示する。」と入力して検索を実行した場合について，処理内容を示す。

ⅰ) 検索ワードが自然文であるため，ゆらぎ辞書，取扱説明書データベースに該当項目が無い。簡易自然文処理(手順6)に移行する。

ⅱ) 簡易自然文処理を実行する。検索ワードから「地デジ」「EPG」「表示」の3個のキーワードが抽出され，手順1へ移行する。

ⅲ) ゆらぎ辞書の中から「地デジ」「EPG」「表示」のゆらぎ語を検索する。その結果，検索ワードの「地デジ」を「地上デジタル」に,「EPG」を「番組表」に変換する。

ⅳ) 取扱説明書データベースから「地上デジタル」「番組表」「表示」が含まれるトピックを検索する。結果，4つのトピックが該当し，検索処理を終了する(手順7)。
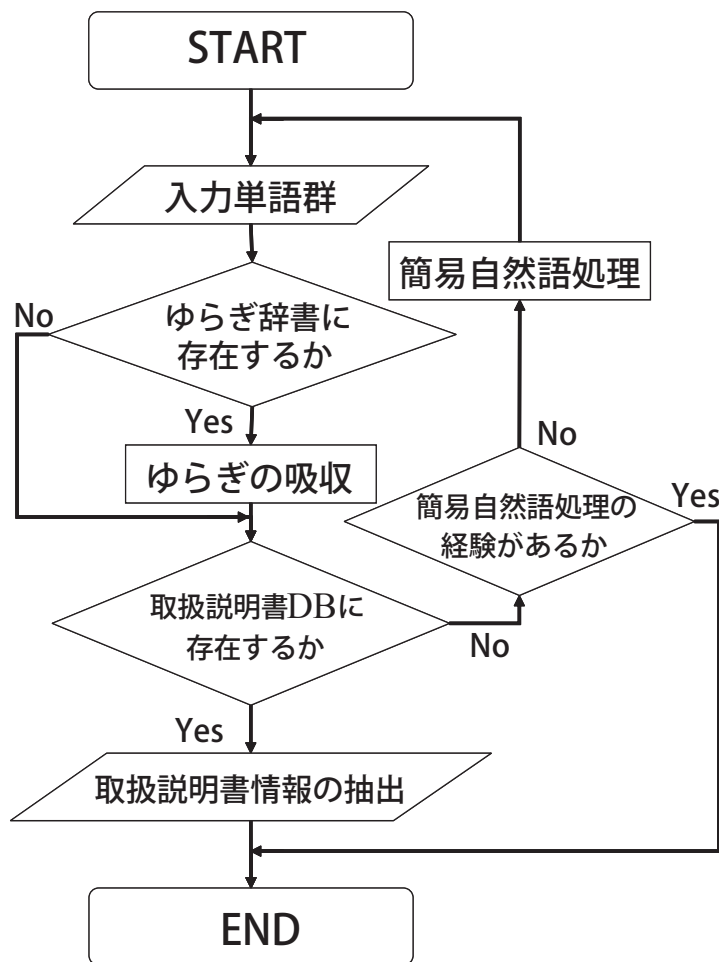
図3　検索アルゴリズム

Fig. 3　Information Retrieval Algorithm

5.3　検索結果のソート

ユーザの検索結果の選択を容易にするために，本システムでは，検索結果をソートして出力している。ソートの指標として採用したのは,各トピックの「重要度」「ヒット数」「目次順」の３種であり，ユーザは検索時に必要に応じてソート方法を選択することが出来る。

5.3.1　重要度ソート

第４章２項で説明した “トピックにおけるキーワードの重要度” に従ってソートする方法である。検索の結果，キーワードがヒットしたトピックごとに重要度データベースから同キーワードの TF-IDF 値を検索し，その値の高い順，つまりユーザが所望する検索結果の順に並び替えを行う。

一方，複数のキーワードによる検索の場合，同じく重要度データベースから，各キーワードの TF-IDF 値を呼び出し，トピック毎に総計を算出する。そして，その総計値の高い順，つまりユーザが所望する検索結果の順に並び替えを行う。

5.3.2　ヒット数ソート

ユーザが過去にアクセスした回数をトピックごとに算出し，それを基準に検索結果をソートする方法である。各項目のアクセス数を,「多くの項目の中から選択された経験」として見なすことで，一般的にニーズが高い情報を優先して表示することが出来る。

5.3.3　目次順ソート

各トピックを，取扱説明書の記述されていた順番に従って並び替える方法である。取扱説明書の記述順位が反映されるため，順を追って取扱説明書の内容を把握することが出来る。

6.　Web 取扱検索システムの概要

我々が開発した知識データベースを応用した Web 取扱検索システム (Web 版検索機能付き取扱説明書) の表示画面を**図 4** に示す。画面右下には，紙冊子と同
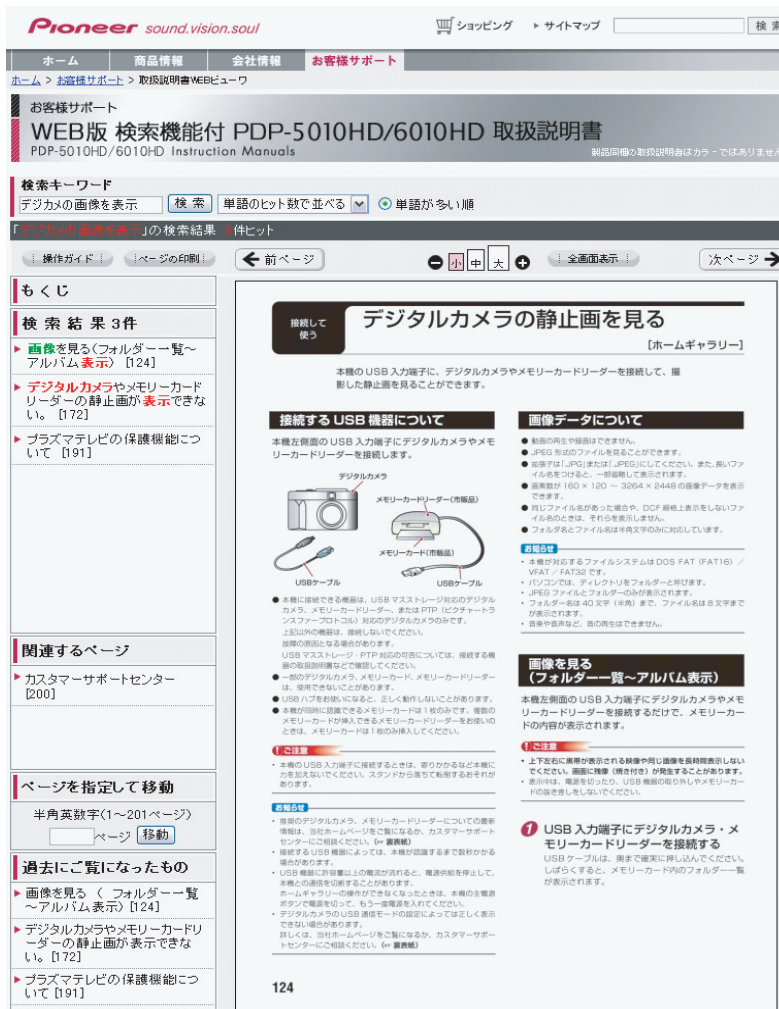
図４　Web 取説画面

Fig .4　Overview of the Web-based Operation Manual

様のページレイアウトの画像が表示されている。その閲覧形態は紙冊子を踏襲しており，ページを進めるのと同様に，次ページの画像へ切り換えることで，読み進めることが出来る。ユーザは，任意に表示ページの切り換え，表示倍率の変更が可能である。

6.1　使用方法

ユーザは，入力フォームに所望する情報のキーワードを入力し，「検索」ボタンをクリックすると検索が開始される。

検索の結果，該当項目のインデックスが左中段の「検索結果」に表示される。インデックスの記述に，ユーザが入力したキーワードが存在する場合，一致箇所を強調して表示される。また同時に，右下に表示されている取扱説明書の画像が，検索結果の最上位の項目が記載されているページへ切り替わる。

このように，ユーザは検索結果のインデックスのリストを選択することによって，該当する項目が記載されている画像を閲覧することが出来る。

7.　まとめ

知識データベースを応用したWeb取扱検索システムを開発し，ユーザが取扱説明書から所望の情報を容易に得られるシステムを実現した。

第一にゆらぎ辞書を利用して，ユーザが入力した検索キーワード(文章の場合には簡易自然文処理で分解可能)を取扱説明書の記述に変換する処理を実行し，ユーザのキーワード入力における負担を軽減できた。

第二に，“トピックにおけるキーワードの重要度”を用いて，検索キーワードに応じたトピックの重要度を算出し，検索結果をユーザにとってより有用と思われる順序で提示することを可能とした。

今後は，本システムを運用することで蓄積される検索履歴，アクセス履歴などのユーザ情報を学習し，知識データベースを自動的に更新する方法についても研究を進めていきたい。

(注1)：
パイオニア製ハイビジョンプラズマテレビ(PDP-5010HDなど)

参 考 文 献

(1) 梶，堀内，莪山：民生機器における知識データベースの構築と応用，PIONEER R&D Vol16，p.62～p.69(2006)

筆 者 紹 介

鎌田　喬浩(かまだ　たかひろ)
技術開発本部 総合研究所システム研究センター 情報メディア技術研究部に所属。主な経歴は，知識データベースの応用技術の研究開発に従事し，現在にいたる。人工知能と認知科学に興味を持つ。

梶　雅代(かじ　まさよ)
技術開発本部 総合研究所システム研究センター 情報メディア技術研究部に所属。主な経歴は，エージェント技術の研究開発，自然言語処理技術の研究に従事し，現在に至る。マルチメディア関連の人工知能応用に興味を持つ。

堀内　直明(ほりうち　なおあき)
技術開発本部 総合研究所システム研究センター 情報メディア技術研究部に所属。主な経歴は，エージェント技術の研究開発に従事し，現在に至る。人工知能とヒューマンマシンインターフェースの研究に興味を持つ。